DOCUMENT RESUME

ED 117 192 95 TM 005 052

AUTHOR	Brown, Bob Burton; Webb, Jeaninne N.
TITLE	The Use of Classroom Observation Techniques in the Evaluation of Educational Programs. TM Report 49.
INSTITUTION	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
REPORT NO	ERIC-TM-49
PUB DATE	Dec 75
CONTRACT	NIE-C-400-75-0015
NOTE	17p.
AVAILABLE FROM	ERIC Clearinghouse on Tests, Measurement, and Evaluation, Educational Testing Service, Princeton, N.J. 08540 (free while supplies last)
EDRS PRICE	MF-$0.76 HC-$1.58 Plus Postage
DESCRIPTORS	*Classroom Observation Techniques; Classroom Research; Elementary Secondary Education; Evaluation Methods; *Program Evaluation; *Student Behavior; *Teacher Behavior

ABSTRACT

Systematic observations of classroom behavior of both teachers and pupils are effective measures of process in education. Until a few years ago, the classroom process--what actually does and does not happen as a result of instituting a new program--went unmeasured for want of tools and techniques to do that job. In this article, examples of observational systems are provided as well as suggestions for tailor making your own observation items for the evaluation of highly unique programs. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED117192

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

TM REPORT 49 DECEMBER 1975

THE USE OF CLASSROOM OBSERVATION TECHNIQUES IN THE EVALUATION OF EDUCATIONAL PROGRAMS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Bob Burton Brown and Jeaninne N. Webb

ABSTRACT

Systematic observations of classroom behavior of both teachers and pupils are effective measures of process in education. Until a few years ago, the classroom process - what actually does and does not happen as a result of instituting a new program - went unmeasured for want of tools and techniques to do that job. In this article, examples of observational systems are provided as well as suggestions for tailor-making your own observation items for the evaluation of highly unique programs.

* * * * *

It has been difficult to find out if an educational program has succeeded or failed and even more difficult to discover why one succeeds or fails. This problem can be attributed to the conceptual blindness which has handicapped evaluation in education for many years.

Types of Measures

In order to think more clearly about evaluation, it will help us to focus on three basic types of measures: 1) presage, 2) process, and 3) product.

Presage means that which forecasts or preindicates. Presage measures tell us about what goes into a program - whether or not we have at hand the resources, facilities, conditions, and people deemed necessary for carrying it out. Accrediting agencies have relied almost entirely upon such presage measures as specifications for physical facilities of the school, size and quality (as determined by degrees and credits) of the faculty, number and range of course offerings, instructional materials, budgetary commitments, and so forth.

TM005 052

The material in this publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Prior to publication, the manuscript was submitted to qualified professionals for critical review and determination of professional competence. This publication has met such standards. Points of view or opinions, however, do not necessarily represent the official view or opinions of either these reviewers or the National Institute of Education.

Process measures describe what happens during the operation of a particular program. Systematic observation of the classroom behavior of both teachers and pupils is an effective measure of process in education. Until just a few years ago, the classroom process went unmeasured for want of tools and techniques to do that job.

Product measures indicate the level of pupil performance that can be attributed to the program being evaluated. Standardized achievement tests are well-known measures of product. However, achievement tests rarely measure the appropriate product or a broad enough range of products. This limitation has been particularly unfortunate for innovative programs that have sought new and unusual products.

Although this paper will stress the importance of process measures, we wish to state that: all three types of measurement -- presage, process, product -- should be included in the evaluation of an educational program. Any thorough evaluation of an educational program should include data regarding 1) the antecedent conditions to 2) the classroom behavior that produces 3) the desired pupil effects.

The current emphasis on accountability in education tends to assess presage or input variables and then skip directly to the measure of product or output without so much as a glance at what happens in between. On the other hand, some enthusiasts for observation of process have shown an inclination for doing their thing to the exclusion of both presage and product measures. Our own earlier work, for example, focused on relationships between teacher beliefs (presage) and teacher practices (process) without giving proper attention to related pupil behavior. Our current work, however, attempts to correct that deficiency.

Systematic observation is more than just a process measure. For example, a teacher's knowledge of and training in the use of observational systems is an important presage variable -- something the teacher brings with him into any educational program in which he participates. Sometimes process is the essential character of the most sought-after outcome of a program, in which case observation becomes the most promising measure of that product. For any program objective which can be stated in behavioral terms, an observational system may be the only appropriate product measure, provided it records pupil behavior as well as teacher behavior in the classroom.

Classroom Observation As a Measure of Achievement

As a sort of "instant" achievement test, systematic observation of classroom teaching-learning processes is an extremely effective technique for the evaluation of educational programs. Collecting pupil-behavior data, pinpointed to the stated objectives of the program, at frequent intervals over a period of several months

provides a most revealing graph of pupil growth. Coupled with similar data regarding what teachers do (and do not do) in the classroom process, observational data lets you zero right in on which aspects of the program are being achieved and which are not in addition to providing insights into changes that might be made in the teaching behavior to produce more desirable results.

Standardized achievement tests -- the paper-and-pencil type taken by pupils sitting in silence at their desks -- rarely measure what innovative programs claim to be their most salient features. It seems ironic that we have been trying to evaluate the educational experiments of the 1960s and 70s with tests conceived a half century ago and infrequently revised only in the smallest details. Such commericially prepared tests designed to appeal to the broadest possible market simply are incapable of measuring what most of us are currently purporting to teach at the local level. Any paper-and-pencil test of pupil performance -- prepared at whatever level for even the most current purposes -- necessarily limits the evaluation to the acquisition of the most superficial informational aspects of the program in question.

Standardized achievement tests are designed to measure global objectives based on traditional programs and seldom provide effective data with which to make decisions about specific programs. The criterion-referenced measurement schemes that now abound on the educational assessment scene are a reaction to the lack of meaningful information available for program evaluation. A specific objective such as "pupils will add two-digit numbers with no carrying" is measured by testing to find out if the pupil can or cannot perform such a task. No norms are involved. It is strictly a "pass-fail" concept. The purpose of the program supposedly is to get every pupil to be able to meet the criterion or standard referred to in this objective. Although such evaluation techniques may be more specifically tailored to measure the objectives of particular programs, they share many of the shortcomings of standardized achievement tests. Usually they are made up of the same old standardized achievement test items, clothed in a new scoring system, which continue to rely heavily on paper-and-pencil responses requiring recall of information or the performance of highly specific skills and tasks. Any measure of this sort by its very nature provides only data <u>after the fact</u> (after the student has been through a particular aspect of the program). Such a measure, no matter how skillfully designed, can yield only information which tells you where the student is -- past the criterion or short of it. How he got there or why he didn't make it remains a mystery, for whose solution there are no data.

An Observation Technique for Evaluating a Parent-Education Program

Throughout the short happy history of this movement, a serious detriment to the imaginative use of systematic observation may well have been the enthusiasm felt by the system's developers and their

devotees for the purity of their system(s), present company included. It was not until about 1970 that we discovered the value of extracting only the most powerful items from well-established systems. Concentration of attention and effort on one of two highly specific items at a time allowed us in the Florida Follow Through Model to train hundreds of parent-educators (usually with no more than a high school education) in the effective use of carefully selected observational items. This experience led to the development of what we call "Desirable Teaching Behaviors."

The Florida Follow Through Model, developed and directed by Ira J. Gordon, is a parent-education program aimed at disadvantaged children, age 4 to 9, and their parents. Parent-educators drawn from disadvantaged areas teach other parents in their neighborhood to participate effectively in the education of their children through educational tasks in the home. Television tapes are made periodically of parents teaching these tasks to their children. The tapes are then analyzed by observers who are trained to look for and record the occurrences of the Seven Desirable Teaching Behaviors, which are listed in Table 1 below.

Table 1

Seven Desirable Teaching Behaviors

Florida Follow Through Model

1. Elicit questions from the learner.

2. Ask questions which have more than one correct answer.

3. Elicit answers of more than one word from learner.
   (a) Encourage learner to enlarge on his response.
   (b) Encourage learner to use complete sentence.

4. Praise learner when he does well.
   (a) Praise learner even when he takes small steps in the right direction.
   (b) Let learner know when he is wrong in a positive or neutral manner.

5. Get learner to evaluate, make judgments or choices on basis of evidence and/or criteria, rather than by random guesses, chance, or luck.

6. Give learner time to think about problem; don't be too quick to help.

7. Give learner time to familiarize himself with task materials. Before starting, give learner introduction or overview.

These are obviously process goals, directed toward improvement of parents' teaching behavior. Not quite so obvious is the fact that most are also linked to pupil or learner behavior. Desirable teaching behaviors should produce desirable behavior on the part of learners, such as asking questions, responding in complete sentences, making judgments on the basis of evidence, and so forth. Incidentally, a test of a good observational item is that it permits reciprocity in that it describes both teacher behavior and learner behavior. When this happens, the item can be used as a "mini" achievement test or product measure.

A teacher is credited with the first Desirable Teaching Behavior (Elicit questions from the learner.) only if the learner is observed to ask a question related to the goals of the task. A teacher is counted as having triggered the fifth Desirable Teaching Behavior (Get learner to evaluate, make judgments, etc.) only if the pupil is observed relating his choice overtly to specified evidence of criteria, such as saying "The marble is small because I can put it in my pocket -- the basketball is large because I can't." By the way, this example also indicated that the third Desirable Teaching Behavior (Elicit answers of more than one word, etc.) has been realized.

A given teaching-learning episode is evaluated by simply counting the number of DTBs recorded or counting the frequency with which each particular DTB is observed. If no DTBs are observed, the program is obviously a failure; if a good number of them are observed in increasing frequency, then we know we are getting someplace. Nothing could be simpler or more powerful in evaluating an educational program.

Data of this type are very useful in analyzing why we may be achieving only partial successes. For example, overemphasis on the fourth DTB (Praise the learner, etc.) may serve to block the realization of DTB 5 (Get learner to evaluate, etc.) if the teacher praises the learner habitually for making lucky guesses. In order to achieve a 5, the teacher may need to change from indiscriminate praising to asking the learner to tell why or how he decided the basketball was big. Any experienced teacher knows that this is not an easy trick to pull off, particularly with shy, frightened youngsters who previously have not been encouraged to engage in an exchange of language with adults. Such observational analyses protect us from the over-achievement of certain goals at the expense of others.

Out of our experience in training paraprofessionals in disadvantaged situtations and the elementary school teachers with

whom they worked in the Florida Follow Through Model also came the idea for the development of the Performance Assessment Record for Teachers (PART), prepared for the Florida Public School Council, Florida Department of Education (3).

Performance Assessment Record for Teachers

The Performance Assessment Record for Teachers (PART) is an instrument for systematically observing, describing, and assessing the classroom performance of teachers and pupils. It is designed primarily for use by teachers in the self-assessment of their own classroom teaching performance. Electronic recordings (either sound or videotape) of classroom behavior serve as the basis of self-assessment procedures. However, this instrument lends itself easily to "live" observations of one's teaching performance by a colleague in what is sometimes called "peer assessment."

The PART does not lend itself well to research designs or to teacher evaluation schemes that ascribe a number or single quantitative description or weight to one's teaching performance. Each of the 22 items on the PART is intended as a word description of what does and does not happen in classrooms. Someone might be interested in keeping score to see how often or how many times these behaviors occur, but frequency scores can be deceiving and this practice is discouraged. Instead, the observer simply should indicate whether or not a given behavior occurred at all during a brief (no more than six minutes) marking period, without differentiating whether it happended once or a dozen times.

None of the 22 items on the PART are intended as universally "good" or "bad" behaviors. Every item has its critics as well as its champions. You cannot help but have your own opinions about the relative merits of the items. Nevertheless, the purpose of the items is not to serve as criteria for deciding who is a good teacher. Instead, they should serve as a framework for looking at and thinking about teaching at more than customary depth -- for the sole purpose of improving the quality of teaching.

The 22 items on the PART in no sense represent all there is to teaching, but only a limited cross section of several possible ways of analyzing teaching-learning situations in classrooms. Such a limitation is necessary if the instrument is to enjoy any practical use by real teachers in real schools.

On the following page is an observation rating scale followed by descriptions of each part of the Performance Assessment Record for Teachers.

PERFORMANCE ASSESSMENT RECORD FOR TEACHERS (PART)

TOTAL	1	2	3	4	5	Teacher Performance Items
						Part I (TPOR)
						1. Some "thing as a thing" (product) made the center of attention.
						2. "Doing something with a thing" (process) made the center of attention.
						3. Learning activity organized around problem or Q posed by T or textbook.
						4. Learning activity organized around problem or Q of genuine concern to pupil.
						Part II (TAX)
						5. P recalls specific information.
						6. P cites ways and means of dealing with specifics.
						7. P uses an abstraction, idea, generalized concept, principle, or theory.
						8. P translates idea from one form to another.
						9. P interprets, gives reason, tells why, shows similarities and differences, cause-and-effect relationships.
						10. P performs directed task or process.
						11. P figures out which previously learned process to apply, then applies.
						12. P shows interaction or relation of elements operating with a process or situation.
						13. P formulates hypothesis (intelligent guess) deliberately.
						14. P devises a classification scheme or evaluative criteria.
						15. P evaluates something on basis of empirical evidence or previously established criteria.
						Part III (TAB)
						16. P openly defends the right of another to hold a value.
						17. P tries to persuade another to accept a value.
						18. P compares and weighs alternative values.
						19. P revises judgments based on evidence.
						Part IV (Soar)
						20. P immobilized through close supervision.
						21. P unrestrained and out of control.
						22. P exercises freedom without the extremes of too much or too little control.

PART I -- (TPOR)

The first four items of this instrument are taken from the Teacher Practices Observation Record (2), which measures the extent to which classsroom activity is inquiry-centered or the extent to which pupils are involved in reflective thinking. These items were selected from the 62 items on the TPOR because experience has shown them to be the most powerful TPOR items in providing teachers with insights and understandings with which to make their teaching more provocative.

PART II -- (TAX)

Items 5 through 15 are derived from the Florida Taxonomy of Cognitive Behavior (14) which is based on Bloom's (1) Taxonomy of Educational Objectives, Cognitive Domain and the modification of that work by Sanders (9). These items measure the level of cognitive or thinking operations used overtly by pupils in the classroom. All items begin with "P" for pupil, signalling that only pupil behavior triggers these items. The teacher is credited for establishing the classroom climate that is conducive to the observed pupil behavior. If an item of behavior occurs, we assume the teacher did something to cause, encourage, or at least permit it to happen. Likewise, if an item fails to be triggered, we assume the teacher did something to prevent, exclude, or circumvent its occurrence.

These items were chosen from the 55-item TAX because they 1) represented the seven major categories of the taxonomy or 2) research and field experience has shown them to be particularly useful in differentiating and influencing classroom performance.

PART III -- (TAB)

Items 16 through 19 are adapted from the Taxonomy of Affective Behavior (9), based on Krathwohl's (6) Taxonomy of Educational Objectives, Affective Domain. These items represent the higher end of the affective scale in which the pupil does much more than play pitch and catch with feelings and sensitivities. The items used here are concerned with valuing, including both the basis for, and process involved in, conceptualizing, organizing, and intellectualizing values. These affective items enjoy very strong ties to the cognitive items in the previous section.

PART IV -- (Soar)

Items 20, 21, 22 stem from the research of Soar (11) who used a variety of observational systems and submitted his findings to factor analysis. These items regarding classroom control represent the factors Soar found that relate most strongly to school achievement on the part of pupils. Items 20 and 21 are both predictive of little or no pupil growth, and item 22 is

characteristic of classrooms in which the greatest pupil growth is made. However, it would be a mistake to leap to the conclusion that items 20 and 21 are "bad" items to be avoided at all costs. Quite the contrary, teachers whose teaching is generally characterized by item 22 frequently make timely use of both items 20 and 21. The question is not which of these items occurs, but whether the teacher limits her classroom control technique to one or both extremes to the exclusion of the highly desirable item 22.

Setting Up an Observational Evaluation

Setting up an observational evaluation system consists of five steps:

1. Select or design an observation system that describes the educational program prior to its implementation,

2. Implement the educational program,

3. Monitor the program while it is in operation,

4. Measure achievement or other anticipated changes,

5. Replicate the program in other settings.

These steps are discussed in the following pages.

Select or Design an Observational System Which Describes the Educational Program: The first step in planning evaluation procedures is to select or design an observation system that clearly describes the educational program to be undertaken, both in terms of what is to occur and what is not to occur. Seldom will a single observation system fully describe a program. Many established systems are available, and you should consult the fifteen-volume Mirrors for Behavior (10) for the most complete description of the available observation techniques. It is usually possible to find two or three observation systems that will describe (or surround) the major features of an educational program. If not, you will have to design one that will.

You can design your own system by selecting appropriate items from established systems, as was done in the PART previously described. Or you can write your own items, as we did for the Desirable Teaching Behaviors also described above. Item writing is not a difficult task. An item must describe behavior of either teacher or pupils (preferably both) that can be observed and recorded in a consistent manner by two or more observers. Two observers are usually all that anyone can afford for the evaluation of programs in the field, which greatly simplifies worries about reliability of the observation items. You count the behaviors on which the two observers agree and throw out those on which they don't. If they disagree more than 10 or 15 percent of the time,

something is wrong; rewrite your items or get yourself a new pair of observers. This may also be a clue that your system of items lacks clarity and easily recognizable relevance to the program (validity). This calls for both a re-think and a rewrite.

The most important criteria for the selection or design of a system is how well it describes the planned program in behavioral terms. No matter how highly pedigreed the system, if it does not measure the salient features of the program, it will provide meaningless information. Examples of the failure to meet this criteria can be found in the countless cases where observational systems utilizing some form of interaction analysis have been used to evaluate programs whose primary goal was to teach students to exhibit problem-solving behavior demanding complex cognitive skills. Interaction analysis (4, 8) is a highly useful technique that will provide measures of many kinds of behaviors but will not provide the slightest clue to the intellectual activity of the classroom.

Implement the Program: Our experience in observing and evaluating educational programs has taught us one alarming fact: Rarely, if ever, is the planned program actually carried out in any recognizable fashion in the classroom. Thus, implementation of the program is not a point to be taken lightly.

In the past two decades, millions of dollars have been poured into the public schools to provide alternatives to traditional education. The evaluation of the effectiveness of these expenditures has been shamefully inadequate. There is a growing awareness of the possibility that all this money and effort may have done more harm than good.

There has been widespread naivete on the part of the administrators of in-service teacher training programs and the consultants they hired to conduct them. It has been assumed that a few days of preschool workshops or a dozen after-school weekly sessions will bring about the desired change in teacher behavior required to implement the funded program. Following participation in these sessions, it is assumed that the teachers involved will go back to the classroom and do right and good forever after, that they will fully implement the new teaching strategies, materials, processes, or interaction patterns planned as the substantive part of the new program. Unfortunately, this is often a false assumption. The teacher, often because of inadequacies in the training provided by the hit-and-run consultant or an inability to transfer what has been learned in the workshop to the classroom, never changes his behavior, and the program is never implemented.

Systematic classroom observation techniques will keep the in-service training activities (and the consultants who design and manage them) honest. By requiring those whose responsibility it is to provide the training for teachers to state in observable terms the behaviors that the teachers (and pupils) are to exhibit, a

careful assessment can be made of the effectiveness of the teacher training as well as the implementation of the new program.

A specific example may illustrate the point. As part of our evaluation of a middle school program that designed a curriculum based on students' concerns, we asked each of the consultants to give us a list of teacher behaviors they intended to increase and a list of those they hoped to decrease as a result of their in-service program. One such behavior all consultants specified was an increase in the use of verbal reinforcement or praise for correct student response. This behavior also reflected the overall objectives of the program and was included as a factor in the observation system designed for the project. It was a simple matter to collect base-line data before the in-service program began by observing the classrooms of the teachers involved and comparing these observation records with those made at the conclusion of the training. Our data clearly indicated that the teachers did change the specified behavior, and significantly so. However, some of the more complex behaviors called for in the program (such as requiring evidence for judgment making and eliciting student application of information) did not occur at all.

If data had been collected at frequent intervals during the training period and fedback to the teachers and in-service consultants, a fuller implementation of the program would have occurred. Our experience indicates that the most effective means of bringing about the desired change in teacher behavior is by a combination of data collecting, feedback to the teacher, and then modeling of appropriate classroom behavior. These activities can be repeated until the teacher can produce the desired behaviors at will and can also elicit specified behavior from the students. Only then can the program designed become the program implemented.

Monitor the Program While It Is In Operation: Classroom observational measurements are even more critical in the on-going assessment of the program than in the final evaluation. Monitoring the program by systematic observation will enable those who are involved in the program to adjust, redesign, or eliminate activities, materials, or strategies found to be inadequate or counter-productive in terms of program goals. This sort of feedback permits the program developers a series of opportunities to tinker with the program until they get it working according to specifications. Data from such observations also provide a graphic history of the success (or failure) of the program, which is invariably more informative than a single statistic obtained from a single-shot evaluation taken only at the conclusion of the evaluation period.

Measure Achievement or Other Anticipated Changes: While the use of standardized achievement tests, attitude or personality measures, and other paper-and-pencil tests is not necessarily precluded by the use of systematic classroom observations, they assume less

importance than usual. If process objectives are to be assessed for which appropriate verbal behaviors are to be the outcome, the only practical measure is systematic observation. Observation data will prove invaluable in interpreting scores obtained by conventional paper-and-pencil measures. For instance, if students in a special reading program fail to make gains on measures of reading vocabulary and observational data show that the pupils had little or no practice in the application of word-attack skills to unknown words, the evaluator has a pretty good clue as to the cause.

Replicate the Program in Other Settings: A major purpose for the evaluation of innovative programs is to help decide whether the program is worth continuing in other settings. The records made by systematic observation can become the guidelines that will insure successful replication of an effective program. Such records show which behaviors, activities, or materials are vital to success and which can be eliminated or reduced to minor roles.

Disadvantages of Classroom Observation Techniques

There are four main disadvantages of classroom observations:

1. Time-consuming
2. Expensive training
3. Administrative burden
4. Threat to teachers

Time Consuming: There is no doubt that classroom observation techniques take time. Twenty- to thirty-minute visits should be made at frequent intervals (at least once a month) throughout the duration of the school year. That means a lot of extra work for someone.

Expensive Training: The training of observers takes time and money. It often requires hiring a consultant to come in and do the training. Released time for the trainees can also run into money.

Administrative Burden: Classroom observation places an increased burden on administrators and supervisory personnel. The responsibility for conventional testing programs usually falls on the shoulders of the classroom teacher and the pupils; systematic observation invariably calls for a third party, someone from the outside. Even if you use peer observers (teachers coming in from next door), you have a problem of arranging for substitutes.

Threat to Teachers: There is no arguing with the fact that initially teachers are threatened by observation. True, this gives away quickly with experience for the really good teachers, but it remains a delicate problem throughout. The threat factor is also

testimony to the power of systematic observation to get at the truth of what does and does not happen in the classroom. This is an excruciating sort of power which may tell and show more than anyone really wants to know. It can be extremely effective or extremely devastating, depending on the skill and care with which it is used. The evaluator who uses systematic observation must be made aware that he is applying a hardnosed technique in a field too often given to tender evasions of reality and responsibility.

A Final Word

Do not use systematic observation techniques unless you are genuinely interested in making something happen. Observation not only evaluates a program, it moves and improves it. It forces or motivates program participants to do what they agreed to do. It does not let any of us get by with lip service and a lick and a promise. It is our view, of course, that the unavoidable effectiveness of systematic observation outweighs the disadvantages we have just mentioned.

In conclusion, permit us to stress the point that evaluation of any educational program is woefully incomplete unless it includes data obtained from a series of systematic observations of the program in operation. Knowing what goes into a program and what comes out of it simply does not constitute an adequate evaluation. That is evaluating everything <u>but</u> the program. There can be no substitute for looking at the program itself -- while it is going on and while something can still be done to make it live up to its promise.

ANNOTATED BIBLIOGRAPHY

1. Bloom, B. S. Taxonomy of educational objectives: Handbook of cognitive domain. New York: David McKay Company, Inc., 1965.

   This classification of educational goals has been a major contribution to educational research and served as the theoretical model for the development of the Florida Taxonomy of Cognitive Behavior.

2. Brown, B. B. The experimental mind in education. New York: Harper & Row, Publishers, 1968.

   As an example of the development of a system to assess both the beliefs and practices of teachers based on a clearly defined educational philosophy, Dewey's Experimentalism, this study is unique in the literature. The Teacher Practices Observation Record, the background of its development, and the complete directions for administration can be found in Chapter 6.

3. Brown, B. B. Performance Assessment Record for Teachers. Institute for Development of Human Resources, Gainesville, Fla.: University of Florida, 1970.

   This is an instrument for systematically observing and recording the classroom performance of teachers and pupils. It is designed primarily for use by teachers in the self-assessment of their own teaching performance. It is an example of an eclectic system.

4. Flanders, N. Teacher influence, pupil attitudes and achievement. Cooperative Research-Monograph Number 12 (Office of Education). Washington, D.C.: Government Printing Office, 1965.

   One of the major influences in the systematic observation movement is this study, Interaction Analyses, a system designed to assess the social and emotional climate of the classroom.

5. Kaplan, L. Taxonomy of affective behavior. Gainesville, Fla.: Institute for Development of Human Resources, University of Florida, 1969.

   The Taxonomy of Affective Behavior is a system designed to assess valuing behavior in the classroom. The items include the basis for the process involved in conceptualizing, organizing, and intellectualizing values. It is based upon Krathwohl's work in classifying the affective domain.

6. Krathwohl, R. Taxonomy of educational objectives. Handbook II: Affective domain. New York: David McKay Company, Inc., 1964.

As a companion to the handbook of the cognitive domain, this classification of educational goals served as the theoretical model for the Taxonomy of Affective Behavior.

7. Medley, D. M. & Mitzel, H. E. A technique for measuring classroom behavior. Journal of Educational Psychology, 1958, 49: 86-92.

The Observation Schedule and Record (OScAR) developed by Medley and Mitzel has undergone modifications and revisions. One of the earlier systems, it assesses several dimensions of the classroom and analyzes both teacher and pupil verbal behaviors.

8. Ober, L., Bentley, L. & Miller, E. Systematic observation of teaching. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1971.

This book fully describes the Reciprocal Category System, a modification of Ned Flanders' Interaction Analysis. Highly specific, the book can be used to train reliable observers in the use of the system.

9. Sanders, N. Classroom questions: What kinds? New York: Harper & Row Publishers, 1966.

A very readable and highly practical application of Bloom's classification of educational objectives aimed at classroom teachers who wish to improve their techniques in assessing student achievement.

10. Simon, A. & Boyer, E. G., eds. Mirrors for behavior. Philadelphia, Pennsylvania: Research for Better Schools, Inc., 1970.

This fifteen-volume anthology of observation instruments is a definitive study of observation systems. It is an invaluable resource to anyone planning evaluation components of educational programs.

11. Soar, R. S. & R. M. An empirical analysis of selected Follow Through programs: An example of a process approach to evaluation, Chapter 11, 1972, National Society for the Study of Education Yearbook, Early Childhood Education, Ira J. Gordon, Editor, 1972.

Robert Soar's research in the relationship between observed classroom behavior of teachers and students and student achievement has made important strides in identifying behaviors which have an impact on student learning.

12. Travers, R. M. W., ed. Second handbook of research on teaching. Chicago: Rand McNally & Company, 1973.

Of particular interest to the reader of this article should be Chapter 5, "The Use of Direct Observation to Study Teaching," by Borok Rosenshine and Norma Furst and Chapter 6, "Techniques of Observing Teaching in Early Childhood and Outcomes of Particular Procedures," by Ira J. Gordon and R. Emile Jester, as well as Chapter 7, "The Assessment of Teacher Competence," by John D. McNeil and W. James Popham.

13. Webb, J. N. & Brown, B. B. Children's concerns observation schedule. Gainesville, Florida: Office of Instructional Resources, University of Florida, 1974.

This observation schedule is an example of a system specifically designed to evaluate an innovative instructional program in the middle school.

14. Webb, J. N. Taxonomy of cognitive behavior: A system for the analysis of intellectual process. Journal of Research and Development in Education. 4: 23-43, 1970.

The Taxonomy of Cognitive Behavior, an instrument designed to measure the cognitive behavior of both students and teachers, is fully described.

70774 • T125P1 5 • 275324 • Printed in U.S.A.